Energy and Policy Considerations for Modern Deep Learning Research

Abstract

The field of artificial intelligence has experienced a dramatic methodological shift towards large neural networks trained on plentiful data. This shift has been fueled by recent advances in hardware and techniques enabling remarkable levels of computation, resulting in impressive advances in AI across many applications. However, the massive computation required to obtain these exciting results is costly both financially, due to the price of specialized hardware and electricity or cloud compute time, and to the environment, as a result of non-renewable energy used to fuel modern tensor processing hardware. In a paper published this year at ACL, we brought this issue to the attention of NLP researchers by quantifying the approximate financial and environmental costs of training and tuning neural network models for NLP (Strubell, Ganesh, and McCallum 2019). In this extended abstract, we briefly summarize our findings in NLP, incorporating updated estimates and broader information from recent related publications, and provide actionable recommendations to reduce costs and improve equity in the machine learning and artificial intelligence community.